Riemannian Optimization for Euclidean Distance Geometry

Chandler Smith (Tufts) Samuel Lichtenberg (Tufts) HanQin Cai (UCF) Abiy Tasissa (Tufts)

Highlights

- Combines non-convex Riemannian matrix completion method and a dual-basis framework
- Comparable reconstruction to state-of-the-art algorithms
- Provable convergence framework

Euclidean Distance Geometry

Euclidean Distance Geometry: Given partial pairwise squared distances $\mathbf{D} = [d_j^i]$ in a matrix, where only some entries are known, can we robustly reconstruct the points $\mathbf{P} = [\mathbf{p}_1 ... \mathbf{p}_n]^T \in \mathbb{R}^{n\times d}$ up to rotation/translation?
Multi-dimensional Scaling (MDS): Recovers $\mathbf{P}$ up to rotation from full information in $\mathbf{D}$ by taking a truncated eigenvalue decomposition of $\mathbf{X} = -\frac{1}{2}(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T)\mathbf{D}(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T)$
Matrix Completion: Algorithms for computing a low-rank matrix $\mathbf{M} \in \mathbb{R}^{n_1\times n_2}$ given a subset of the entries $\Omega = \{(i,j) \in [n_1] \times [n_2] \mid M_{ij} \text{ is known}\}$. Original methods[1] developed were convex minimizations of the nuclear norm

$$\min_{\mathbf{X}\in\mathbb{R}^{n_1\times n_2}} \|\mathbf{X}\|_* \text{ subject to } \mathcal{P}_\Omega(\mathbf{X}) = \mathcal{P}_\Omega(\mathbf{M})$$

where $\mathcal{P}_\Omega$ is defined as

$$\mathcal{P}_\Omega(\cdot) = \sum_{(i,j)\in\Omega} \langle \cdot, \mathbf{E}_{ij}\rangle \mathbf{E}_{ij}$$

for $\mathbf{E}_{ij} = \mathbf{e}_i\mathbf{e}_j^T$. Many scalable non-convex algorithms for this problem exist.
Problems with existing methods: Distance matrices are a difficult set to optimize over due to the triangle inequality. This leads to poor recovery results with standard matrix completion algorithms on $\mathbf{D}$.

Existing Work and Geometric Structure

Riemannian Methods for Matrix Completion [2]: A non-convex Riemannian approach to matrix completion.

- **Main idea:** Non-convex gradient descent scheme for matrix completion using entries.
- **Formulation:** Wei et al. define the following optimization program

$$\min_{\mathbf{X}\in\mathbb{R}^{n\times n}} \langle \mathbf{X} - \mathbf{M}, \mathcal{P}_\Omega(\mathbf{X} - \mathbf{M})\rangle \text{ subject to } \operatorname{rank}(\mathbf{X}) = r$$

The algorithm is a gradient descent scheme on the manifold of rank r matrices with a tangent space at the l-th iterate $\mathbf{X}_l = \mathbf{U}_l\mathbf{\Sigma}_l\mathbf{V}_l^T$ defined as $\mathbb{T}_l = \{\mathbf{U}_l\mathbf{Z}_1^T + \mathbf{Z}_2\mathbf{V}_l | \mathbf{Z}_1, \mathbf{Z}_2 \in \mathbb{R}^{n\times r}\}$. To update to $\mathbf{X}_{l+1}$, the update is taken in the gradient descent direction projected onto the manifold $\mathbb{T}_l$, then retracted back to the rank r manifold. More specifically

$$\mathbf{X}_{l+1} = \mathtt{SVD}_r(\mathbf{X}_l + \eta_l \mathcal{P}_{\mathbb{T}_l}\mathcal{P}_\Omega(\mathbf{M} - \mathbf{X}_l))$$

with $\mathtt{SVD}_r$ defined as the truncated SVD of rank r and η_l computed through an exact line search

- **Pros:** Proven convergence results, efficient implementation
- **Cons:** Poor recovery for the EDG problem

Dual Basis Approach

Idea: Instead of optimizing over distance matrices, move to Gram matrices for easier computability.

Constructing Dual Basis and Sampling Operator: Following [3], accessible information is in the form of

$$D_{ij} = \|\mathbf{p}_i - \mathbf{p}_j\|_2^2 = \|\mathbf{p}_i\|^2 + \|\mathbf{p}_j\|^2 - 2\mathbf{p}_i^T\mathbf{p}_j = X_{ii} + X_{jj} - 2X_{ij}$$

Defining $\mathbf{w}_\alpha = \mathbf{E}_{\alpha_1,\alpha_1} + \mathbf{E}_{\alpha_2,\alpha_2} - \mathbf{E}_{\alpha_1,\alpha_2} - \mathbf{E}_{\alpha_2,\alpha_1}$ for $\alpha = (\alpha_1, \alpha_2)$, we can represent accessible information as $\langle \mathbf{X}, \mathbf{w}_\alpha\rangle$. Given this new basis and its Gram matrix $\mathbf{H}$ defined by $\mathbf{H}_{\alpha,\beta} = \langle \mathbf{w}_\alpha, \mathbf{w}_\beta\rangle$, the dual or bi-orthogonal basis can be constructed as

$$\mathbf{v}_\alpha = \sum_\beta \mathbf{H}_{\alpha,\beta}^{-1}\mathbf{w}_\beta$$

This allows us to define an analogous sampling operator for the dual basis problem:

$$\mathcal{R}_\Omega(\cdot) := \frac{L}{m}\sum_{\alpha\in\Omega} \langle \cdot, \mathbf{w}_\alpha\rangle \mathbf{v}_\alpha$$

This problem defined on Gram matrices is mathematically equivalent to standard matrix completion on the squared distance matrix, although as $\mathcal{R}_\Omega$ is not self-adjoint we consider a computable surrogate instead.
Defining Computable Surrogate and Optimization Program: We construct a computable surrogate and its corresponding objective function as follows:

$$\mathcal{R}_\Omega^\star\mathcal{R}_\Omega(\cdot) := \frac{L^2}{m^2}\sum_{\alpha,\beta} \langle \cdot, \mathbf{w}_\alpha\rangle\langle \mathbf{v}_\alpha, \mathbf{v}_\beta\rangle \mathbf{w}_\beta$$

$$\min_{\mathbf{X}\in\mathbb{R}^{n_1\times n_2}} \langle \mathbf{X} - \mathbf{M}, \mathcal{R}_\Omega^\star\mathcal{R}_\Omega(\mathbf{X} - \mathbf{M})\rangle \text{ subject to } \operatorname{rank}(\mathbf{X}) = r$$

RieEDG

Algorithm: Fusing the dual-basis approach with the efficient Riemannian scheme presented in [2].

- **Main idea:** Define a similar algorithm as in [2], but substituting our computable surrogate operator $\mathcal{R}_\Omega^\star\mathcal{R}_\Omega$
- **Pros:** Provable convergence framework given good enough initialization
- **Cons:** Slower time complexity with similar reconstruction results as other non-convex algorithms [3]

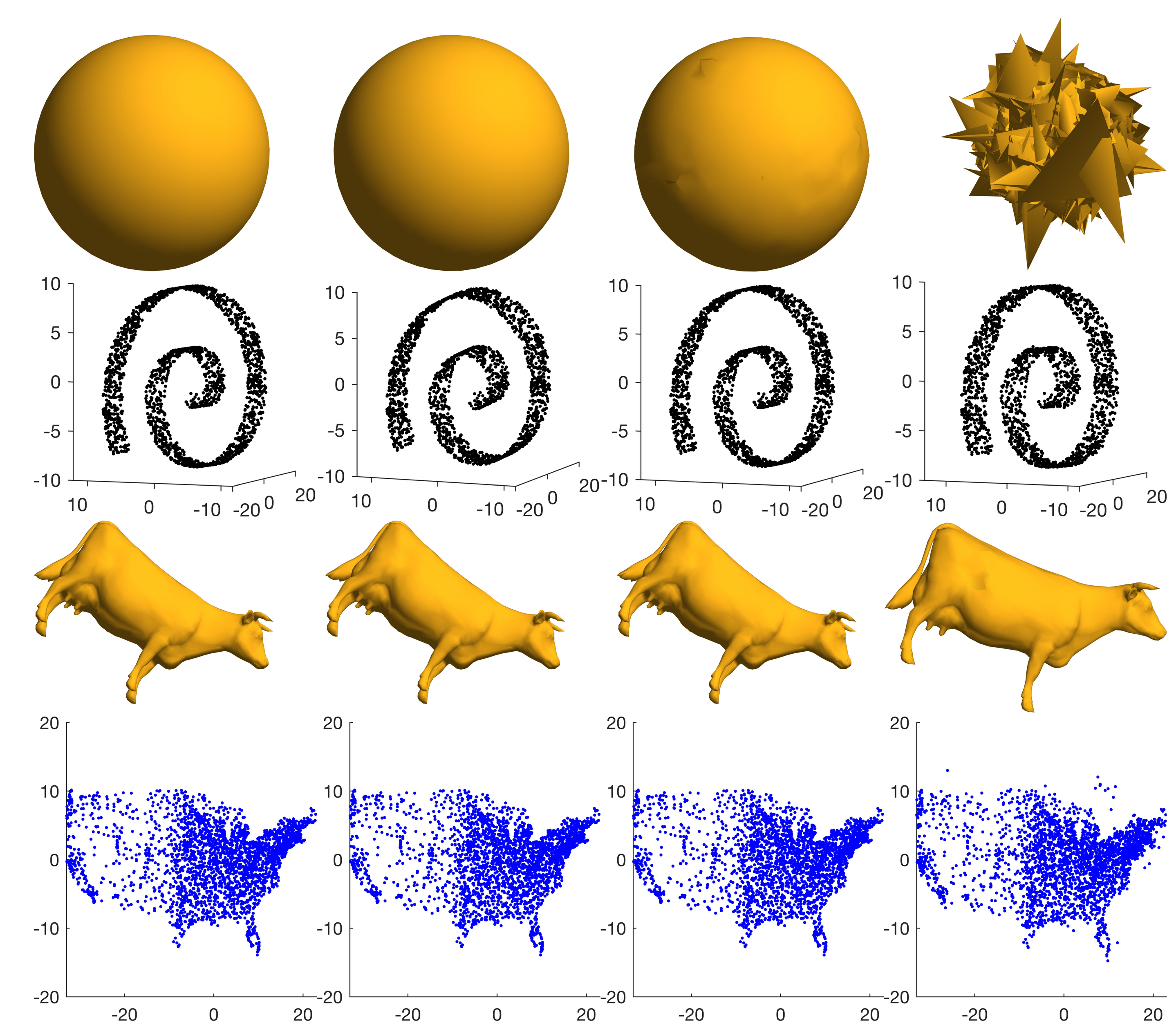
Algorithm: RieEDG
Input: $\mathcal{P}_\Omega(\mathbf{D})$: The observed distance information; k: the dimension of the datapoints; η: the step size
Initialize $\mathbf{X}_0 = \mathtt{EVD}_k(\mathcal{R}_\Omega^\star\mathcal{R}_\Omega(\mathbf{X})) = \mathbf{U}_0\mathbf{\Lambda}_0\mathbf{U}_0^T$
for $l = 0, 1, 2 \cdots$ **do**
$\mathbf{G}_l = \mathcal{R}_\Omega^\star\mathcal{R}_\Omega(\mathbf{X} - \mathbf{X}_l)$
$\mathbf{W}_l = X_l + \eta\mathcal{P}_{\mathbb{T}_l}\mathbf{G}_l$
$\mathbf{X}_{l+1} = \mathtt{EVD}_k(\mathbf{W}_l)$
end for
Output: $\mathbf{X}_{\text{rev}}$

Numerical Experiments

Synthetic data and Tabulated Results: Various 2- and 3-dimensional datasets were used for testing and are referred to below in increasing size order. The objective of RieEDG is to recover the full set of points $\mathbf{P}$ up to orthogonal transformation from a subset of entries of $\mathbf{D}$ chosen using a Bernoulli sampling model, where each entry has a probability γ of being selected for $\gamma \in [0, 1]$, with an expected γL entries chosen. RieEDG outputs the Gram matrix $\mathbf{X} = \mathbf{P}\mathbf{P}^T$, from which $\mathbf{P}$ can be recovered. The comparison referenced in Table is the relative error between the recovered matrix $\mathbf{X}_{\text{rev}}$ and the ground truth matrix $\mathbf{X}$ in Frobenius norm averaged over 10 trials. Each run was terminated after 500 iterations or a relative difference of 10^{-7} in Frobenius norm.

Dataset	$\gamma = 5\%$	3%	2%	1%	5% Timing (sec)
Sphere (3D)	6.2e-07	1.2e-06	9.52e-03	1.08	4.62
Swiss Roll (3D)	5.04e-07	8.84e-07	1.14e-06	0.0604	30.9
Cow (3D)	5.58e-07	8.62e-06	1.50e-06	0.0095	67.4
U.S. Cities (2D)	5.90e-07	1.613-03	0.0168	0.0796	135

Reconstructed Images: Below are images of the reconstructed datasets. From left to right, the sampling rate goes from 5% to 1% as in the table above.



References

[1] B. Recht, "A simpler approach to matrix completion," *Journal of Machine Learning Research*, 2011.

[2] K. Wei, J.-F. Cai, T. F. Chan, and S. Leung, "Guarantees of riemannian optimization for low rank matrix completion," *Inverse Problems and Imaging*, 2020, ISSN: 1930-8337. DOI: 10.3934/ipi.2020011.

[3] A. Tasissa and R. Lai, "Exact reconstruction of euclidean distance geometry problem using low-rank matrix completion," *IEEE Transactions on Information Theory*, 2019. DOI: 10.1109/TIT.2018.2881749.

Acknowledgements

This work is partially supported by NSF DMS 2208392 and NSF DMS 2304489.